CSTD - Commission on Science and Technology for Development

Regulation of AI in Scientific Research

Overview

The integration of Artificial Intelligence (AI) into scientific research represents one of the most significant shifts in the history of human discovery, comparable to the invention of the microscope or the peer-review system. We are witnessing a transition from "hypothesis-driven" science where human intuition guides experimentation to "data-driven" science, where machine learning (ML) algorithms analyze vast datasets to identify patterns invisible to the human eye. In 2024, this transformation is no longer theoretical; it is operational across every major discipline. From DeepMind's AlphaFold predicting protein structures[1] to Google's GNoME discovering millions of new materials,[2,3] AI is compressing decades of research into weeks.

However, this rapid acceleration has outpaced the development of necessary regulatory frameworks. The scientific community is now grappling with a "reproducibility crisis" exacerbated by proprietary "black box" algorithms; systems where the internal decision-making process is opaque even to their creators. Furthermore, the deployment of these tools raises profound ethical questions regarding data privacy, algorithmic bias in medical research,[4,5] and the

---

1 "The Revolutionary Impact of AlphaFold on Drug Discovery." Lindus Health, https://www.lindushealth.com/blog/the-revolutionary-impact-of-alphafold-on-drug-discovery-decoding-the-mystery-of-protein-folding.
2 "Google DeepMind's materials AI has already discovered 2.2 million new crystals." Tanaka Precious Metals, 2024. https://tanaka-preciousmetals.com/en/elements/news-cred-20240110/.
3 "Millions of new materials discovered with deep learning." Google DeepMind, 2023. https://deepmind.google/blog/millions-of-new-materials-discovered-with-deep-learning/.
4 "Artificial Intelligence Bias in Health Care: Web-Based Survey." JMIR Medical Education, https://pmc.ncbi.nlm.nih.gov/articles/PMC10337406/.
5 "Real-world examples of healthcare AI bias." Paubox, 2024. https://www.paubox.com/blog/real-world-examples-of-healthcare-ai-bias.

potential for dual-use technologies (such as AI designing novel pathogens).[6,7,8]

In 2025, the World Bank and UNCTAD report that the "Scientific Digital Divide" is widening. While a handful of nations and private corporations consolidate control over the most powerful AI models, the Global South risks becoming a passive consumer of scientific knowledge rather than an active producer. The challenge for the Commission on Science and Technology for Development (CSTD) is to foster an environment where AI accelerates innovation without compromising the integrity, safety, and equity of the scientific enterprise.

What is the CSTD?

The UN Commission on Science and Technology for Development (CSTD) is a subsidiary body of the Economic and Social Council (ECOSOC). Established in 1992, it serves as the United Nations' focal point for science, technology, and innovation (STI) for development.[9] The Commission's primary mandate is to analyze how STI can contribute to achieving the 2030 Agenda for Sustainable Development (SDGs), specifically focusing on the interface between science policy and development outcomes.

Unlike the Security Council, the CSTD is not a binding enforcement body; rather, it acts as a global forum for strategic planning, the facilitation of high-level discussions, and the sharing of "lessons learned" among member states.[9] It is also responsible for the periodic review of the

---

6 "Artificial intelligence challenges in the face of biological threats: emerging catastrophic risks for public health." Frontiers in Artificial Intelligence, 2024. https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1382356/full.
7 "AI and the Evolution of Biological National Security Risks." CNAS, https://www.cnas.org/publications/reports/ai-and-the-evolution-of-biological-national-security-risks.
8 "Dual-use capabilities of concern of biological AI models." PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12061118/.
9 "About the CSTD." UNCTAD, https://unctad.org/topic/commission-on-science-and-technology-for-development/about.

World Summit on the Information Society (WSIS) outcomes. In the context of Artificial Intelligence, the CSTD plays a critical normative role. It identifies emerging trends and advises the UN General Assembly on how to regulate technologies that have transboundary impacts. For this session, the CSTD wishes to examine potential safeguards to regulate AI in scientific research, ensuring that the scientific method remains robust, transparent, and universally accessible in the algorithmic age.

The Paradigm Shift: Pre-AI vs. Post-AI Research

To understand the necessity of regulation, one must distinguish between the traditional scientific method and the emerging AI-driven paradigm.

Pre-AI Research (Traditional Science): historically, scientific discovery was a linear, human-centric process. A researcher would observe a phenomenon, formulate a hypothesis, design an experiment to test it, collect data, and analyze the results using standard statistical methods. This process was inherently limited by human cognitive capacity. In other words, a scientist can only read so many papers and analyze so many variables. Collaboration was often slow, and verification relied on other scientists physically repeating experiments.

Post-AI Research (In Silico Discovery): the current era is defined by high-throughput data analysis and "in silico" (computer-simulated) experimentation. In this model, AI algorithms process massive datasets, genomic sequences, climate records, material properties to generate hypotheses and predict outcomes before a physical experiment is ever conducted.

- **Generative Discovery:** Instead of testing materials one by one, AI models can hallucinate plausible chemical structures and simulate their stability.

- **Automation**: "Self-driving labs" now exist where AI algorithms direct robotic arms to mix chemicals, analyze results, and plan the next experiment without human intervention.[10]

While this exponentially increases efficiency, it fundamentally changes the nature of scientific verification. If the "scientist" is a proprietary algorithm owned by a private corporation, how can the broader community verify the results?

Case Studies in AI-Driven Science

1. AlphaFold and the Protein Revolution

Perhaps the most famous example of AI in science is AlphaFold, developed by Google DeepMind. For 50 years, the "protein folding problem", predicting a protein's 3D structure from its 1D amino acid sequence. This was a grand challenge in biology. In 2020, AlphaFold solved this, and by 2024, the AlphaFold Protein Structure Database had expanded to include over 200 million predicted structures, covering nearly all known proteins.[1]

- **Impact:** This has revolutionized drug discovery, allowing researchers to skip months of expensive X-ray crystallography work.[1]
- **Regulatory Concern:** While the database is open, the most advanced versions of such models are increasingly proprietary. If drug discovery relies on a tool that only one company understands, global health security becomes dependent on that single private entity.

[10] "Recent Breakthrough in AI-Driven Materials Science." arXiv, 2024. https://arxiv.org/pdf/2402.05799.

2. GNoME and Material Science

In late 2023, DeepMind and Berkeley Lab unveiled the Graph Networks for Materials Exploration (GNoME). This AI tool discovered 2.2 million new crystals, including 380,000 stable materials that could power future batteries, solar panels, and superconductors.[2,3]

- **The "A-Lab":** To validate these predictions, an autonomous laboratory (A-Lab) synthesized 41 of these new materials in 17 days, a rate completely unachievable by human chemists.[10]
- **Regulatory Concern:** The rapid discovery of dual-use materials. The same algorithms that find better battery materials could theoretically identify potent neurotoxins or unstable energetic materials (explosives), lowering the barrier for bad actors to discover dangerous compounds.[7,8]

Key Points of Concern

Transparency and the "Black Box" Problem: A core tenet of science is reproducibility; one scientist should be able to repeat another's experiment and get the same result. However, deep learning models are often "black boxes." They contain billions of parameters, and the specific path the algorithm took to reach a conclusion is often uninterpretable. If a researcher publishes a paper based on an AI prediction but cannot explain why the AI made that prediction, is it science? This "interpretability gap" threatens to undermine trust in the scientific record.

Bias Mitigation in Medical Research: AI models are not objective; they are a reflection of the data they are trained on. Historical scientific data contains deep-seated biases.[4]

- **Medical Devices:** Recent audits found that AI-driven pulse oximeters and skin cancer detection algorithms performed poorly on patients with darker skin tones because the training data was overwhelmingly Caucasian.[5]
- **Kidney Function:** Algorithms predicting acute kidney injury were found to be less accurate for women because the training datasets contained only 6% female representation.[5]
- **The Risk:** If these biased models become the "standard" for scientific research, they will codify health disparities into the infrastructure of modern medicine, leading to misdiagnoses and unequal treatment on a global scale.[4]

Data Privacy and Consent: Training robust AI systems requires massive amounts of data. In fields like genomics and social science, this data is derived from human subjects. There is a growing controversy regarding "broad consent." Did a patient who donated tissue for a specific cancer study in 2010 consent to have their genetic sequence used to train a commercial AI model in 2025? The commercialization of scientific data by third-party AI companies creates a murky legal environment regarding ownership and privacy.

Dual-Use and Biosecurity: The "democratization" of science via AI has a dark side. In 2022, researchers famously tweaked a drug-discovery AI (originally designed to find non-toxic therapies) to instead search for toxic molecules. In less than six hours, the AI designed 40,000 potentially lethal chemical warfare agents, including VX nerve gas.[6,7,8] The CSTD must address how to regulate the release of open-source AI models that could lower the barrier to entry for creating biological or chemical weapons.

The Global Landscape and The Digital Divide

The development of AI in science is not evenly distributed, leading to a profound "Scientific Digital Divide." The major hubs of AI development are concentrated in the Global North: the United States, China, the United Kingdom, Canada, and the European Union.

The Infrastructure Gap: according to UNCTAD, fewer than a third of developing countries have national AI strategies, and 118 nations are largely absent from global AI governance discussions.[9] The "compute divide" is equally stark; training a model like AlphaFold requires millions of dollars in cloud computing resources, which are inaccessible to most universities in the Global South.

The New Colonialism: there is a risk of "scientific colonialism," where developing nations provide the raw data (e.g., genetic diversity, climate data) while the value-added analysis and intellectual property (IP) are captured by entities in the Global North. If the tools for advanced research become proprietary services charged by the hour, the Global South may be permanently priced out of cutting-edge discovery.

Current Regulatory Landscape

The European Union (EU AI Act): The EU AI Act is the world's first comprehensive AI law. Notably, it includes a "research exemption" for AI systems developed solely for scientific R&D, attempting to avoid stifling innovation. However, it imposes strict rules on "high-risk" AI systems (e.g., medical devices).[11] The tension for the CSTD is defining where "research" ends and "product" begins.

United States & China: the US has focused on "voluntary commitments" from major AI labs and Executive Orders restricting the export of high-end AI chips to rival nations. China has implemented regulations on "Deep Synthesis" technologies and generative AI, focusing on social stability and state control.

UNESCO Recommendation on the Ethics of AI: adopted by 193 states, this recommendation emphasizes that AI should function as a public good. It explicitly calls for international cooperation to share AI infrastructure and data for scientific research, a principle the CSTD seeks to operationalize.

Proposed Solutions to Regulate AI in Scientific Research

To address these challenges, the CSTD considers the following mechanisms for regulation, as outlined by recent policy discussions:

- **Mandatory Disclosure of AI Use:** Establishing protocols where researchers must disclose the use of AI tools in all stages of research. This includes citing specific models and the origin of training data to ensure reproducibility.
- **Expansion of Ethics Review Boards:** Institutional Review Boards (IRBs) currently

[11] "High-level summary of the AI Act." EU Artificial Intelligence Act, https://artificialintelligenceact.eu/high-level-summary/.

review human subject research. Their mandate should be expanded to include AI expertise, allowing them to assess the ethical implications of algorithms before research begins.

- **Licensing and Certification:** Implementing a licensing or certification framework for "science-grade" AI tools. This ensures that AI systems used in regulated research meet minimum standards of accuracy and safety.
- **Risk-Based Government Oversight:** Adopting frameworks similar to the EU AI Act, where AI systems are categorized by risk. High-risk systems (e.g., those affecting patient health) would require strict government oversight, while lower-risk tools would face fewer barriers.
- **AI Governance Committees and Audits:** Creating specialized AI governance committees within research institutions to oversee compliance. This includes mandating **routine audits** of datasets and models to check for bias, accuracy, and reproducibility.
- **Researcher Training:** Implementing required training programs for scientists in AI ethics and transparency, ensuring they understand the limitations and risks of the tools they employ.
- **Watermarking and Digital Provenance:** Enforcing the use of digital watermarking and provenance tracking to clearly distinguish between AI-generated data and human-generated experimental data.
- **Open Access to Data and Models:** Promoting open science by requiring that models and code used in publicly funded research be made accessible to the global scientific community.

- **Responsible Research Norms:** Encouraging professional societies and journals to establish and enforce norms for responsible AI use in publication and peer review.
- **Interdisciplinary Collaboration:** Fostering collaboration between ethicists, scientists, and technologists to ensure that technological advancement aligns with human values.
- **Strict Regulation of Dual-Use Research:** Implementing strict regulations and monitoring for dual-use AI research, specifically regarding biosecurity and chemical synthesis, to prevent the creation of harmful agents.

Questions to Consider

1. **Transparency vs. IP:** How can we ensure the transparency and reproducibility of AI-driven discoveries without violating the intellectual property rights of the private companies developing these tools?
2. **Ethical Frameworks:** What specific ethical frameworks should guide the use of AI in research involving human subjects, specifically regarding consent for data use in training sets?
3. **Human Oversight:** What role should traditional peer review play in an era of AI? Should there be a "human-in-the-loop" requirement for all AI-generated scientific conclusions?
4. **Bias Management:** How do we standardize the auditing of datasets to manage the risk of "AI bias" in scientific conclusions and innovation?
5. **Global Equity:** How can international standards be established that protect safety but do not stifle innovation in developing countries with varying levels of technological infrastructure?

Works Cited

"Artificial intelligence challenges in the face of biological threats: emerging catastrophic risks for public health." *Frontiers in Artificial Intelligence*, 2024. https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1382356/full.

"About the CSTD." *UNCTAD*, https://unctad.org/topic/commission-on-science-and-technology-for-development/about.

"AI and the Evolution of Biological National Security Risks." *CNAS*, https://www.cnas.org/publications/reports/ai-and-the-evolution-of-biological-national-security-risks.

"Artificial Intelligence Bias in Health Care: Web-Based Survey." *JMIR Medical Education*, https://pmc.ncbi.nlm.nih.gov/articles/PMC10337406/.

"Dual-use capabilities of concern of biological AI models." *PMC*, https://pmc.ncbi.nlm.nih.gov/articles/PMC12061118/.

"Google DeepMind's materials AI has already discovered 2.2 million new crystals." *Tanaka Precious Metals*, 2024. https://tanaka-preciousmetals.com/en/elements/news-cred-20240110/.

"High-level summary of the AI Act." *EU Artificial Intelligence Act*, https://artificialintelligenceact.eu/high-level-summary/.

"Millions of new materials discovered with deep learning." *Google DeepMind*, 2023. https://deepmind.google/blog/millions-of-new-materials-discovered-with-deep-learning/.

"Real-world examples of healthcare AI bias." *Paubox*, 2024.

https://www.paubox.com/blog/real-world-examples-of-healthcare-ai-bias.

"Recent Breakthrough in AI-Driven Materials Science." *arXiv*, 2024. https://arxiv.org/pdf/2402.05799.

"The Revolutionary Impact of AlphaFold on Drug Discovery." *Lindus Health*, https://www.lindushealth.com/blog/the-revolutionary-impact-of-alphafold-on-drug-discovery-decoding-the-mystery-of-protein-folding.